УДК 004.89:37.041

О.С. Вороной, Г.А. Єгошина
Державний університет інформатики і штучного інтелекту, м. Донецьк, Україна
alex.voronoy@gmail.com

Засоби інтеграції онтологій предметних областей для створення баз знань інтелектуальних навчальних систем

У статті розглядається організація системи формування і оновлення онтологічної бази знань інтелектуальної навчальної системи. Запропоновано алгоритм і склад програмного забезпечення інструментальних засобів об'єднання онтологій предметних областей.

Вступ

У даний час інтенсивно розвивається напрямок використання технологій глобальної мережі Інтернет у системі освіти. Зручність і гнучкість гіпертекстового представлення матеріалів, оперативний доступ до великих обсягів інформації, можливість віддаленого спілкування між викладачами та студентами дозволяють широко застосовувати інтернет-технології при створенні систем дистанційного навчання. Однак існуючі системи не забезпечують наявність смислових зв'язків між знаннями навчальних курсів і є просто бібліотекою статичних гіпертекстових підручників і тестових завдань, чого явно недостатньо для повноцінної та ефективної організації навчального процесу. Це виключає можливість автоматизованого аналізу знань програмними засобами і обмежує інтелектуальні можливості навчальної системи з побудови індивідуальних планів навчання. Таким чином, актуальною є задача створення інтелектуальних навчальних систем (ІНС), заснованих на систематизованих знаннях.

Ефективним засобом представлення та систематизації знань є онтології. Онтології використовуються для формальної специфікації понять і відносин, які характеризують певну галузь знань. Перевагою онтологій як способу представлення знань є їх формальна структура, яка спрощує комп'ютерну обробку.

ІНС, що використовують онтології, знаходять застосування при розробці та моделюванні систем представлення та інтерпретації інформації в процесі навчання. Онтологія визначає терміни, за допомогою яких можна описати і структурувати предметну область. Використання онтології ефективне під час пошуку і об'єднання інформації з різних джерел і середовищ. Мова онтології використовується для надання інформації чітко визначеного значення і являє собою загальний набір термінів для опису та подання предметної області, що вивчається. Будучи аналогом поняття «модель», онтологія служить засобом комунікації між учнем і викладачем.

У даний час на базі проведеної консорціумом W3 реорганізації сервісу Web мережі Інтернет в Semantic Web розвивається два напрями використання web-онтологій у навчальному процесі. Перший напрямок пов'язаний зі створенням, наповненням та оновленням баз знань ІНС на основі онтологій. Другий – з побудовою тестуючих програмних систем нового покоління, що оперують знаннями, які містяться в онтологіях.

Для першого напрямку актуальне рішення задачі автоматизації поетапного створення онтології предметної області.

Постановка завдання

Завдання об'єднання онтологій, що виникло практично з початком використання онтологій в інформаційних технологіях, продовжує залишатися актуальним через неможливість використання уніфікованих онтологій в специфічних і динамічних галузях знань.

Онтологія складається з термінів (понять), їх визначень і атрибутів, а також пов'язаних з ними аксіом і правил виводу. Детальне формальне визначення онтології в вигляді кортежу з семи елементів наведено в [1]. У рамках даної роботи будемо використовувати більш спрощене визначення, запропоноване Т.А. Гавриловою та Ф.В. Хорошевським в [2].

Згідно з цим визначенням формально онтологія представляється трійкою:

$$O = < T, R, F >, \quad (1)$$

де T – концепти (терміни) предметної області (ПО), яку описує онтологія O;

R – відношення між термінами ПО;

F – функції інтерпретації, задані на термінах і/чи відношеннях онтології O.

Завдання об'єднання онтологій формулюється таким чином: дано дві правильні онтології, потрібно створити третю правильну онтологію, яка представляє поняття у вхідних онтологіях, а також додаткові обмеження і взаємозв'язки, якщо вони потрібні.

Дві онтології O1 і O2, створені різними розробниками для однієї предметної області, будуть суттєво відрізнятися і за таксономією, і за лексикою, і деякою мірою за семантикою. Семантична неоднорідність онтологій є наслідком опису однакових онтологічних концептів різними термінами (синоніміка), призначення різних значень одному і тому ж концепту в різних контекстах (багатозначність). Неоднорідність в онтологіях може також викликатися різними таксономіями концептів.

З цієї причини задача об'єднання нетривіальна і в основному може вирішуватися в напівавтоматичному режимі, тому що на поточному рівні розвитку онтологічних визначень тільки експерт може остаточно підтверджувати коректність семантики встановлених відносин між онтологічними поняттями.

Отже, існує велика потреба у високоефективних засобах автоматизації порівняння та інтеграції онтологій.

У задачах порівняння та об'єднання онтологій при поетапному створенні бази знань ІОС існує чотири основні проблеми.

1. Семантичні відмінності: деякі онтологічні концепти описують однаковий домен різною термінологією (синонімія і багатозначність). Це призводить до доменів, що перетинаються. Таким чином, існує потреба в засобах картування, які змогли б інтерпретувати описи навчальних ресурсів, використовуючи лексичний та семантичний аналіз для того, щоб вирішити проблеми синонімії та багатозначності.

2. Структурні відмінності: різні розробники онтологій анотують цифрові навчальні ресурси різними онтологічними концептами. Це створює синтаксичні відмінності, вимагаючи інструментальні засоби для об'єднання, які змогли б різні таксономії об'єднувати в загальну таксономію.

3. Масштабування операцій картування та об'єднання онтологій: це важливо особливо для великих онтологічних репозиторіїв.

4. Відсутність апріорних знань: апріорні знання необхідні для картування та об'єднання онтологій з використанням контрольованих методів видобутку знань. Однак такі знання не завжди доступні. У цьому випадку неконтрольовані методи необхідні як альтернатива при відсутності апріорних знань.

Існуючі інструментальні засоби інтеграції онтологій

Інструменти об'єднання онтологій допомагають користувачам знайти подібність і відмінність між вхідними онтологіями і створюють результуючу онтологію, яка містить елементи вхідних онтологій. Для досягнення цієї мети вони автоматично визначають відповідності між концептами у вхідних онтологіях або забезпечують середовище, де користувач може легко знайти і визначити ці відповідності. Ці інструменти відомі як інструменти відображення, вирівнювання та об'єднання онтологій.

Відображення (mapping) онтології полягає в знаходженні семантичних зв'язків подібних елементів з різних онтологій. Вирівнювання (alignment) онтологій полягає в тому, щоб встановити різні види відповідності (або зв'язку) між двома онтологіями, а потім повторно зберегти вхідні онтології і таким чином надалі використовувати інформацію одне одного. Об'єднання (merging) онтологій – генерація однієї узгодженої онтології з двох вхідних.

Дослідники різних галузей інформатики працюють над автоматичним або підтримуваним інструментально об'єднанням онтологій, однак ці дослідження та розробки знаходяться на ранніх стадіях і потребують подальшого розвитку.

Існуючі інструментальні засоби, які мають справу з находженням відповідності між онтологіями та їх об'єднанням, можна поділити на чотири групи:

1) для об'єднання двох онтологій з метою створення однієї нової (PROMPT, Chimaera, OntoMerge);

2) для визначення функції перетворення з однієї онтології в іншу (OntoMorph);

3) для визначення відображення між концептами у двох онтологіях, знаходячи пари відповідних концептів (наприклад, OBSERVER, FCA-Merge);

4) для визначення правил відображення для зв'язку тільки релевантних частин вхідних онтологій (ONION).

Великий перелік та опис існуючих у світовій практиці програмних систем та інструментальних засобів для роботи з онтологіями наведено в [3]. Загальним для названих інструментальних засобів є наявність інтерактивних редакторів онтологій, які в різний спосіб керують діями користувача в процесі перетворення онтологій.

При цьому з деякими особливостями реалізації в усіх системах або створюються списки пропонованих операцій (PROMPT), або вказуються місця в онтології, де потрібне втручання експерта (Chimaera), або визначається набір операторів перетворення, які можна застосувати до онтології (OntoMorph), або знаходяться і вказуються користувачу синоніми у вхідних онтологіях (OBSERVER).

В OntoMerge нова онтологія є об'єднанням двох вхідних онтологій і набору аксіом з'єднання. Процес об'єднання полягає в трансляції обох онтологій до загального синтаксичного вигляду на розробленій авторами мові. При створенні об'єднаної онтології використовуються аксіоми з'єднання, машина виведення та інструменти для трансляції екземплярів даних вхідних онтологій в об'єднану.

У FCA-Merge використовуються математичні методи, щоб створити сітку концептів, що зв'язує концепти вхідних онтологій. Алгоритм використовує відносини еквівалентності і підклас-суперклас. Система ONION заснована на алгебрі онтології. Використовуються лексичні методи і методи на основі графів.

Метод знаходження лексичної подібності між іменами концептів використовує словники та методи семантичної індексації, засновані на місцезнаходженні групи слів у тексті.

Організація системи формування та оновлення знань інтелектуальної навчальної системи

У рамках проекту багатоагентного інтелектуального навчального середовища для циклу дисциплін напряму підготовки студентів «Комп'ютерні науки» авторами розробляються засоби формування і оновлення знань ІОС на основі неструктурованих джерел [4] і семантичного Web [5]. Основні компоненти системи формування онтологічної бази знань представлені на рис. 1.

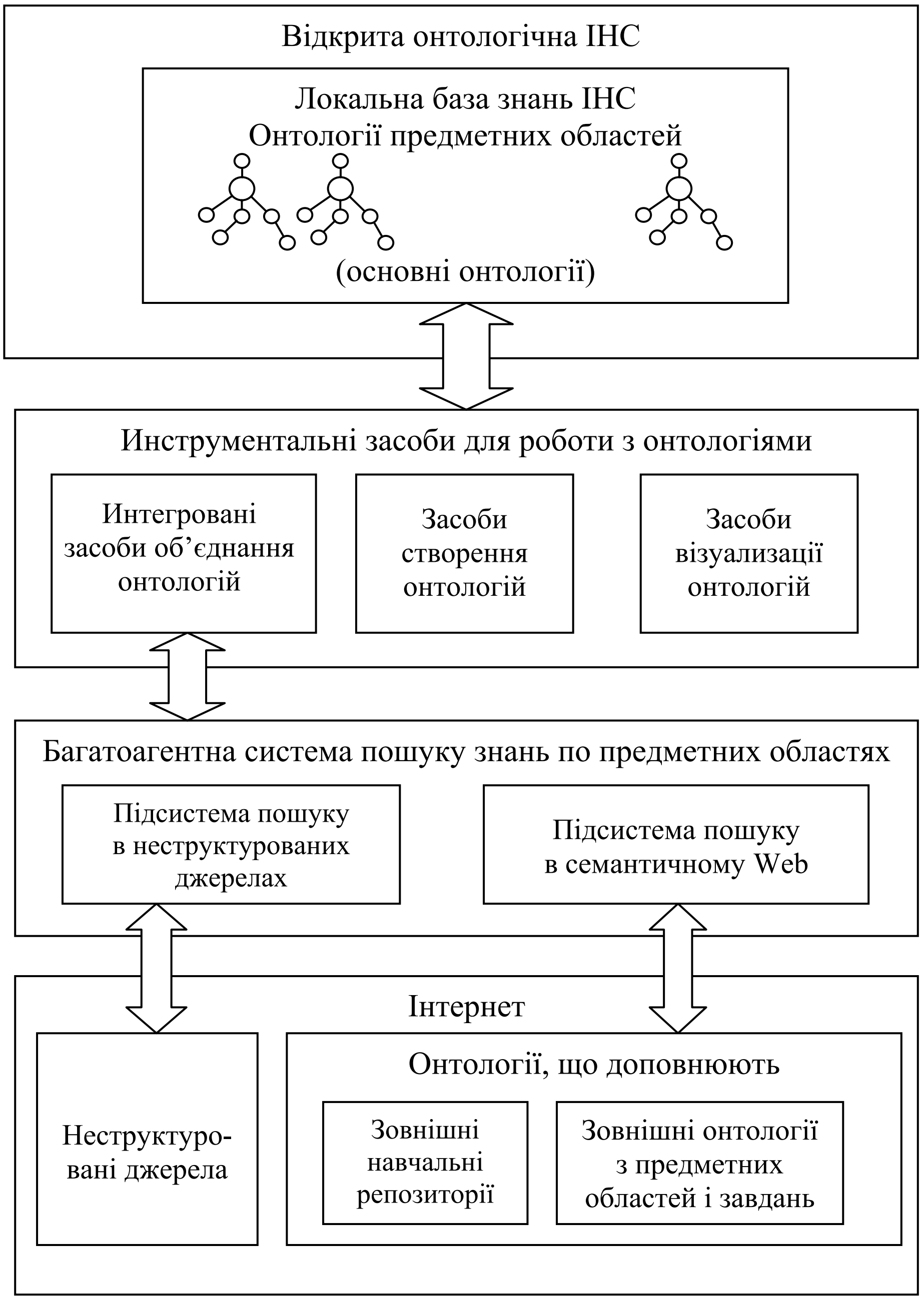


Рисунок 1 – Система формування знань онтологічної ІНС

Розробка онтології у конкретній предметній області знань являє собою досить складний і тривалий процес, коли протягом декількох років група розробників формує набір понять предметної області онтології. Тому система формування онтологічної бази

знань містить інтегровані засоби об'єднання онтологій, що дозволяють створювати і доповнювати основні онтології предметних областей навчального середовища шляхом використання предметних знань (частини знань ПО), представлених в семантичному Web. Такі онтології будемо називати онтологіями, що доповнюють, або для спрощення – складовими.

Джерелом складових онтологій є багатоагентна система пошуку онтологічної інформації з предметних областей. Пошукові запити в цій системі формуються на основі тезауруса ПО та завдань експерта локальної бази знань. Знайдені релевантні складові онтології із зовнішніх навчальних репозиторіїв та інших ресурсів семантичного Web завантажуються пошуковою системою в робочу базу даних інтегрованих засобів об'єднання онтологій.

Розробка алгоритму і програмних засобів інтеграції онтологій

Інтегровані засоби об'єднання онтологій, що пропонуються авторами, включають чотири підсистеми: відображення, вирівнювання, порівняння та об'єднання онтологій. Завдання об'єднання онтологій є найбільш часто вирішуваним при створенні бази знань ІОС. Тому розглянемо більш детально роботу підсистеми об'єднання онтологій. Онтологія розглядається як граф класів. Кожне поняття є вершиною, а відносини клас-підклас є дугами. Клас без базового класу є вершиною графу. Дозволено множинне спадкування. Будемо розглядати онтології, що представляються єдиним графом понять з однією кореневою вершиною. Граф не має посилань на зовнішні онтології.

Алгоритм об'єднання онтологій включає наступні укрупнені дії.

1. Вирівнювання порядків онтологій.
2. Розв'язання конфлікту імен.
3. Об'єднання класів та їх властивостей і відносин.

Під порядком онтології розуміється кількість рівнів ієрархії графу. Пропонований алгоритм може працювати при виконанні двох вимог, які пред'являються до онтологій, що об'єднуються:

– кількість рівнів ієрархії повинна бути однаковою;

– не повинно бути збігу імен понять, що знаходяться в онтологіях на різних рівнях ієрархії, тобто повинен бути відсутнім конфлікт імен.

В алгоритмі одна з онтологій вважається базовою, інша – тою, що доповнює. Алгоритм буде працювати більш ефективно, якщо базова онтологія більше, ніж та, що доповнює. Імовірність того, що онтології, які об'єднуються, будуть мати різну ієрархічну структуру, досить велика. Для коректної роботи запропонованого алгоритму необхідно вирівнювання порядків онтологій. Вирівнювання порядків – це процес приведення ієрархічної структури онтології з меншою кількістю рівнів ієрархії до структури другої онтології. Процес реалізується шляхом додавання в ієрархічну структуру фіктивних рівнів ієрархії. Імена понять на фіктивних рівнях ієрархії не розглядаються при роботі алгоритму.

Сутність пропонованого алгоритму полягає в наступному.

Крок 1. Формування списку збіжних імен понять онтологій, що знаходяться на різних рівнях ієрархії.

Крок 2. Формування списку збіжних імен понять онтологій, що знаходяться на однакових рівнях ієрархії.

Крок 3. Формування списку імен понять онтології, що доповнює, для яких відсутня синонімічність понять у базовій онтології на однакових рівнях ієрархії.

Список, отриманий на кроці 1 до початку об'єднання онтологій, має бути розглянутий експертом з метою розв'язання існуючого конфлікту імен. Тобто конфлікт імен повинен бути усунений до роботи алгоритму на кроці 2 і 3. На кроці 2 синонімічні поняття онтологій аналізуються модулем «перевірка відповідності». На кроці 3 незбіжні поняття обробляються модулем «аналіз елементів, що не збігаються».

Склад та взаємодія модулів підсистеми представлені на рис. 2. Розглянемо функції модулів.

Модуль визначення порядків онтології. У даному модулі проводиться визначення порядку вхідних онтологій шляхом присвоєння кожному елементу властивості «рівень». Порядок онтології визначається елементом з максимальним рівнем. У модулі реалізується сортування елементів онтологій за значеннями рівнів ієрархії, що дозволяє підвищити ефективність роботи алгоритму. Якщо порядки онтологій не збігаються, то необхідне їх вирівнювання.

Модуль вирівнювання порядків онтологій. Для онтології з меншим порядком додаються фіктивні класи в кількості, що дорівнює різниці порядків онтологій.

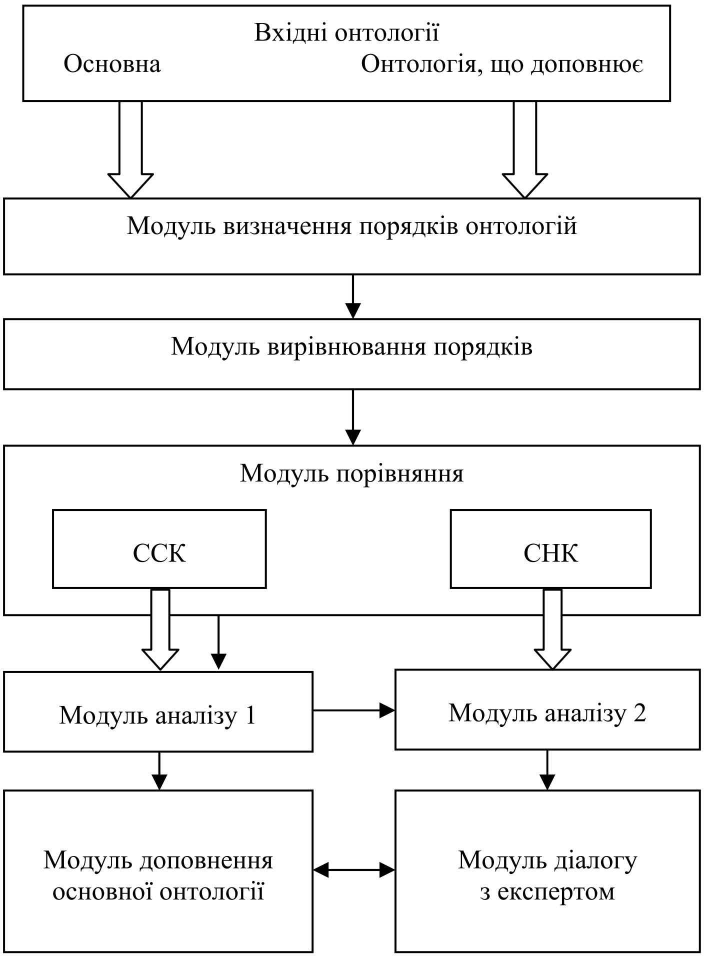


Рисунок 2 – Склад і зв'язки модулів підсистеми об'єднання онтологій

Модуль порівняння. У цьому модулі проводиться порівняння елементів онтології, що доповнює, з усіма елементами базової онтології, що знаходяться на однаковому рівні. В результаті роботи формуються два списки:

1) список збіжних імен концептів (ССК);

2) список концептів, що не збігаються (СНК), який містить елементи онтології, що доповнює, відсутні в базовій онтології.

Модуль «Перевірка відповідності» (модуль аналізу 1). Робить аналіз кандидатів на синонімічність у списку ССК. Для кожного елемента зі списку здійснюється порівняння базових класів і підкласів в основній онтології і онтології, що доповнює. Порівняння підкласів здійснюється за мажоритарним принципом (збігу за більшістю). У разі повного збігу, який означає, що поточні елементи в онтологіях повністю ідентичні, виконується перехід до наступного елементу списку. У разі мажоритарного збігу відсутні підкласи з онтології, що доповнює, додаються в основну. У випадку розбіжності, що означає суперечність: наявність в об'єктів з однаковими іменами різних базових класів і різних підкласів, видається повідомлення експерту для вирішення суперечності. Елементи, що додаються в базову онтологію за результатами мажоритарного збігу, повинні бути виключені зі списку СНК.

Модуль «Аналіз елементів, що не збігаються» (модуль аналізу 2). На етапі роботи цього модуля в списку РНК залишаються елементи, які відсутні в основній онтології, але присутні в онтології, що доповнює. Це означає присутність нової гілки в таксономії онтології, що доповнює, яку необхідно приєднати до певного ієрархічного рівня базової онтології. Це завдання покладається на експерта і може бути успішно виконане у разі відсутності порушень ієрархічної структури гілки, що доповнює. Якщо кожний наступний елемент у списку РНК має рівень, що дорівнює або на одиницю менший попереднього, то порушень в ієрархічній структурі немає.

Висновки

Запропонований алгоритм та програмне забезпечення для об'єднання онтологій є важливою складовою частиною інструментальних засобів інтеграції інформаційних ресурсів для створення онтологічних баз знань. Використання розроблених засобів дозволять прискорити створення інтелектуального навчального середовища для підготовки студентів за напрямком «Комп'ютерні науки». Завданням подальших розробок і досліджень має бути реалізація інших складових системи інтеграції, формалізація процесів інтеграції та розробка більш універсальних алгоритмів розширення онтологій ПО.

Література

1. Maedche A. Clastering Ontology – based Metadata in the Semantic Web / A. Maedche, V. Zacharias // Proceedings PKKD-2002. – Helsinky, 2002. – P. 342-360.
2. Гаврилова Т.А. Базы знаний интеллектуальных систем / Т.А. Гаврилова, В.Ф. Хорошевский. – СПб. : Питер, 2001.
3. Тузовский А.Ф. Системы управления знаниями (методы и технологии) / Тузовский А.Ф., Чириков С.В., Ямпольский В.З. ; под общ. ред. В.З. Ямпольского. – Томск : Изд-во НТЛ, 2005. – 260 с.
4. Грунский И.С. Агентно-ориентированная система извлечения знаний из распределенных источников информации / И.С. Грунский, С.М. Вороной // Материалы X Международной научно-технической конференции «Актуальные проблемы информационно-компьютерных технологий, мехатроники и робототехники», (Крым, Кацивели, 28 сентября – 3 октября 2009 г.). – Т. 2. – С. 364-366.
5. Вороной А.С. Використання онтологій для підвищення якості пошуку інформації для поповнення баз знань інтелектуальних систем / А.С. Вороной // Матеріали Міжнародної науково-технічної конференції «Комп'ютерні науки і інженерія 2009», (Львів, 14 – 16 травня 2009 р.). – Львів, 2009. – С. 364-366.

А.С. Вороной, А.А. Егошина

Средства интеграции онтологий предметных областей для баз знаний интеллектуальных обучающих систем

В работе рассмотрена организация системы формирования и обновления онтологической базы знаний интеллектуальной обучающей системы. Предложен алгоритм и состав программного обеспечения инструментальных средств автоматизации объединения онтологий предметных областей.

A.S. Voronoj, A.A. Yegoshina

Tools for Integration Domain Ontologies for Knowledge Bases of Intelligent Tutoring Systems

In this paper we describe the formation and updating of the intelligent tutoring system, that uses ontological knowledge base. The algorithm and the content of software automation tools for integration domain ontologies are proposed.

Стаття надійшла до редакції 06.04.2010.